DOCUMENT RESUME

ED 338 253 IR 053 780

AUTHOR Langschied, Linda; And Others
TITLE Machine-Readable Data Files at Rutgers: A Preliminary Report of the Task Force on Numeric Data in Machine-Readable Form.
INSTITUTION Rutgers, The State Univ., New Brunswick, N.J. Univ. Libraries.
PUB DATE 91
NOTE 28p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Access to Information; Cataloging; College Libraries; Computers; *Databases; Higher Education; *Library Collection Development; Library Education; Library Networks; Library Planning; *Library Services; *Numeric Databases; *Policy Formation
IDENTIFIERS *Rutgers the State University New Brunswick NJ

ABSTRACT

The Task Force on Numeric Data in Machine-Readable Form at Rutgers University Libraries was formed in February 1991. The group was made up of librarians representing all of the Rutgers campuses and areas of library services such as public services, collection development, and technical services. The charge to the task force was to: (1) recommend collection development policies related to numeric data in machine-readable form; (2) recommend appropriate levels of service; (3) determine the training and skills needed to provide service; (4) recommend policies on access and hardware; (5) recommend policies on cataloging; and (6) recommend a plan for implementing these recommendations. The task force's recommendations in these areas are included in this paper, as are certain key recommendations which are highlighted at the outset of the report. These key recommendations include creating a Machine-Readable Data Files (MRDF) Committee, forming a university-wide network to be known as RUNet, providing immediate action to address Government Printing Office (GPO) data in the libraries, and investigating the possibility of centralizing data services. Appended to the paper are the collection development policy statement for MRDF, a collection profile of MRDF, and descriptions and names of scientific data sources. (MAB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

MACHINE-READABLE DATA FILES AT RUTGERS:

A PRELIMINARY REPORT OF THE

TASK FORCE ON NUMERIC DATA IN MACHINE-READABLE FORM

ED338253

Task Force Members:

Linda Langschied, Chair, Alexander Library
Ka Neng Au, Dana Library
Mary Jane Cedar Face, Computing Services, User Services
Howard Dess, Library of Science and Medicine
Mary Beth Fecko, Library Technical and Automated Services
Jim Nettleman, Paul Robeson Library
Michele Ruhlin, Alexander Library
Jane Sloan, Douglass Library
Robert Sewell, Library Administration, ex officio

Submitted May 1, 1991

120537780

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Linda Langschied

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

TABLE OF CONTENTS

Introduction.....i

I. Service Aspects of Machine-Readable Data Files.....1

A. An Overview.....1
B. Levels of Service.....3
C. Immediate Concerns.....5

II. Collection Development Issues.....6

A. An Overview.....6
B. Outline of a Selection Process.....7

III. Cataloging Machine-Readable Data Files.....9

A. An Overview.....9
B. Materials to be Cataloged: Library-Owned.....10
C. Materials to be Cataloged: Non-Library Owned.....11
D. The Bibliographic Record.....11

IV. Access to Machine-Readable Data Files.....12

A. An Overview.....12
B. The Current Situation at Rutgers.....13
C. Some Technical Problems.....13

V. Training Issues.....14

A. Staff Training.....14
B. User Training.....15

Appendices

INTRODUCTION

Machine-readable data files (MRDF) on campus are a proliferating body of information. Once stored primarily on magnetic tape, and housed at university computer centers, they were a relatively manageable set of informational resources. Currently, the amount of machine-readable data, particularly numeric data, is growing unabated, and to complicate matters, is being distributed in a wide variety of computer formats. Data are released in various media -- round tape, cartridge, CD-ROM, and diskette -- and the hardware needed to handle these formats is as varied. Data itself may be served up "raw," or come packaged with front-end software, or may be accessed with unrelated, commercial software. Obviously, providing service to MRDF has become a very complicated business. The challenge to provide responsible access services to MRDF requires a an approach that cuts across traditional boundaries between Library Services (LS) and Computing Services (CS). Data service acknowledges the marriage of information and technology; the complementary skills of both the library and computing sides of Information Services are needed to mount a successful MRDF service at Rutgers.

In February, 1991, the Associate University Librarian for Research and Undergraduate Services formed the Task Force on Numeric Data in Machine-Readable Form, a group comprised of librarians representing all of the Rutgers campuses and areas of library service: public services, collection development and technical services. Additionally, membership included a representative from Computing Services' User Services division. The Task Force's charge was to:

* recommend collection development policies related to numeric data in machine-readable form.

* recommend levels of service

* determine training and skills needed to provide service

* recommend policies on access and hardware

* recommend policies on cataloging

* recommend a plan for implementing recommendations

One key addition to this original charge was the decision to include full-text, non-bibliographic data files, along with numeric data files, within the purview of the Task Force.

To address these issues, the members of the Task Force divided into subcommittees to work on specific areas of concern, and to author portions of the report. Several Task Force members volunteered to serve as subcommittee coordinators, responsible for the organization of workload and for collating input. Subcommittee assignments were arranged as follows:

Service Issues:
Linda Langschied, Coordinator
Jim Nettleman
Mary Jane Cedar Face

Collection Development:
Howard Dess, Coordinator
Mary Jane Cedar Face
Michele Ruhlin
Jane Sloan
Bob Sewell

Cataloging Policies:
Mary Beth Fecko, Coordinator
Jim Nettleman
Michele Ruhlin

Access Issues:
Ka Neng Au, Coordinator
Mary Jane Cedar Face
Mary Beth Fecko
Linda Langschied

Training Issues:
Linda Langschied, Coordinator
Ka Neng Au
Jane Sloan

The Task Force met three times to discuss philosophy and policy for data services, and communicated electronically on an ongoing basis. At the Task Force's final meeting, it was decided that several of our key recommendations should be highlighted at the outset of the report. Those recommendations are as follows:

1. Creation of the MRDF Coordinating Committee

The establishment of an "MRDF Coordinating Committee" as a permanent advisory group appears to have considerable merit for continuing the work begun by the Task Force. It should be comprised of individuals from different parts of the Library system and Computing Services, as well as campus data users, to review and discuss the broadest conceivable range of MRDF issues. Moreover, as a Coordinating Agency within the New Jersey State Data Center, we are contractually required to have a database advisory committee that meets regularly with its constituents to determine their data needs. The committee that once provided this service is now defunct; the MRDF Coordination Committee could fulfill this obligation to the State Data Center.

2. Completion of the University-wide Network (RUNet)

This is the single most important technical priority for establishing equitable access to data resources among the RU campuses.

3. Immediate Action to Address GPO Data in the Libraries

The data dissemination practices of the GPO present an immediate crisis for the libraries. In particular, the libraries are becoming overwhelmed with a proliferation of numeric data on compact disk. A committee of LS and CS personnel should be immediately created to address this urgent situation

4. Investigate Possibility of Centralizing Data Services

Centralization of services for MRDF will eliminate wasteful duplication of resources, both material and human, and will help ensure equitable access. Creation of the position of MRDF Bibliographer or Coordinator should be considered. Again, a committee should be formed to pursue the feasibility of reorganizing our current system of data services provision.

I. SERVICE ASPECTS OF MACHINE READABLE DATA FILES

A. AN OVERVIEW

1. Service Must Correlate with Demand

> The successful academic library is demand-driven, not supply oriented. It begins with the specific scholarly/information needs of its clients, not the speculative acquisition and warehousing of a broad range of resources. Past library use studies indicate that a significant percentage of material acquired in research libraries is seldom if ever used. The needs of the scholar will differ by discipline and within disciplines; the library must be attuned to these differences, and will ensure a balance in collections and access which parallels the research and instruction carried out by the academic programs.
>
> (University of Alberta Library Self Study Report: Riding the Wave, October 1990.

The general statement above may be easily extended to provision of data services at Rutgers. While data files in electronic format are increasingly used by students and faculty, their specialized nature requires that we provide services in a judicious manner. Active liaison with campus computer file users, particularly the faculty, must be achieved in order for us to determine actual need.

However, this stance should not infer that we should be passive collectors or providers of service. A balance must be arrived at between meeting demand (a reactive stance) and anticipating demand (a proactive stance). The reactive stance may be appropriate in some instances because:

1. The technology changes extremely rapidly.
2. The technology, services, and some data are costly.
3. The materials and services may be out of the "mainstream," or used by a small fraction of the academic community.
4. Provision of service for MRDF is complex and often requires specialized knowledge of hardware and software.

However, in other instances, a proactive stance becomes essential and there is a responsibility to anticipate demand in the provision of services, because:

1. When new services are provided in anticipation of demand, use is usually made of these services.
2. With a reactive stance, we will be unprepared to meet the challenge of the proliferation of MRDF and changing technologies.
3. While MRDF may currently be used by only a small fraction of the academic community, this fraction may be research intensive. In other words, service provision for MRDF is a necessary part of research support in a university environment.

Recommendation:

There needs to be a balance between a proactive and reactive stance towards MRDF. The basis for determining the appropriateness of servicefor MRDF may vary according the characteristics of the particular item in question. Cost, ease of use (or lack thereof), ensuring collection integrity -- all will effect decisions on MRDF collecting and service levels. The MRDF Coordination should establish a clear-cut set of service criteria to ensure consistency and coordination in the decision-making process.

2. Local Resources Will Never be Adequate.

The time has come for libraries to abandon the notion that they can independently provide collections and services to meet all campus research needs. Clinging to the belief that "bigger is better" can only serve to dissipate our buying power in an era of standstill library budgets. A new institutional paradigm is needed.

Resource sharing is a concept that has been in place for years, yet only minimally implemented. It requires commitment to the concept that one's own institutional emphasis on ownership must change to one of access. Fortunately, in the area of machine-readable files, resource sharing is currently the norm: Rutgers participates in data consortia like ICPSR, and the New Jersey State Data Center, for much of its data. However, as data becomes available from more sources, in more formats, and through different vehicles, eg. electronic bulletin boards, we must carefully consider what we choose to add to the collections. Additionally, special consideration must be given to commercially-produced materials, which may be costly to duplicate among units.

An area of concern, which may be beyond our ability to affect, but which must be addressed, is the acquisition of data all across campus. We know that several departments and Institutes at Rutgers--eg, Bureau of Government Research, Bureau of Economic Research, the Chemistry and Geology departments--are in the habit of procuring their own data. We should establish a means of knowing what data is available outside of LS and CS, and whether it is available for use by other members of the Rutgers community.

Recommendations:

There are ways that limited resources can be more effectively utilized. It is essential that Information Services develop networking, coordination of services, and other mechanisms for promoting cost-effective services, resource sharing, and access to MRDF

We urge that high priority be given to the completion of the university-wide network: this is the single most important technical component for ensuring timely and equitable access to computer files at Rutgers.

A campus-wide inventory of machine-readable data files is also recommended.

B. LEVELS OF SERVICE

1. Service Provision Must Be Well-Defined, Yet Flexible

It is important to define areas of responsibility between Library and Computing Services, and among the Library units. The aim should be to avoid duplication of effort as much as possible, yet be flexible enough to allow units to determine the appropriate level of service locally. This implies a clear statement of each level of service, and the locus for each level. Obviously, a team approach is needed to determine the divisions of responsibility. The creation of the MRDF Coordinating Committee will allow us to begin to build a long-range, overall structure of service.

2. Setting a Minimum Threshold of Service

Independent of long-term planning, we can and should immediately establish a basic service threshold which should be available at all LS locations. All units should be able to provide fundamental reference/referral/advisory services to MRDF as they would to any other library materials, regardless of format. Much of our service pattern follows directly from the types of material we are expected to service. Numeric, applications and instructional MRDF are not now clearly the responsibility of LS, and may more properly belong in CS: this is an area in which clear delineation of responsibility must yet be developed. Bibliographic files, however, with which we should and must be expert, clearly fall within LS. Skills necessary to provide this minimum level of service include:

--Ability to recognize the information needs that are most appropriately satisfied by numeric databases.
--Ability to identity Rutgers data holdings and supporting documentation through the OPAC.
--Awareness of bibliographic finding aids to identify appropriate data, whether owned by RU or not, eg. catalogs of government data collections, ICPSR Guide, codebooks, technical documentation.
--Knowledge of online databases that act as "bibliographic" locators for MRDF, eg. POLL, CD-NET.
--Ability of subject specialists to generally include MRDF within their purview.
--Ability to identify the "referral point" -- when a user needs to seek more expert help. Librarians should act as mediators.

This level of service should be as non-dependent on any one individual as possible. Instead, organizational structure and mechanisms need to be clearly defined, thoroughly understood, and relied upon in setting the minimum service thresholds.

3. Beyond the Minimum Threshold

The next level should speak to the appropriateness of the item identified to the needs and abilities of the potential user. This may well be subject dependent, and therefore assigned to specific LS locations dependent on subject. It may also be hardware or software dependent, and thus fall more in CS territory. Possible skills needed:

--Knowledge of the characteristics of numeric databases, information they contain, and typical uses of the databases themselves
--Ability to assist in selection of appropriate media
--Knowledge of storage media and machine compatibilities

The highest service level addresses the actual access to, processing of, and interpretation of output. It is this level that clearly requires integration of traditional CS and LS strengths. It is also this level which may be so costly on a per use basis that it can be afforded (if at all) only at one location in the University. Should this be the case, remote access to this service should be the norm, to promote equity across the entire University. Possible services provided at this level:

--Assisting patrons with software for accessing data.
--Helping users extract data.
--Provide "convenience search" information: obtain a fact from a machine-readable file.
--Creating databases and extracting, manipulating or reformatting data; creating data sub-sets.

4. Service Levels and Intrastructure

Currently, delivery of service to MRDF at Rutgers involves a system that has arisen with little planning: history, computing strengths of certain individuals, and local needs are all factors in the development of our current infrastructure. The system which was appropriate ten years ago--the division of informational responsibilities between CS and LS, based on physical format of information sources--is no longer viable. Multiple formats of the same data have necessitated the involvement of both LS and CS in data services. Our roles as information providers have come to overlap, a truism made concrete by the recent restructuring of the libraries and computing centers under one administrative umbrella.

Recognizing this reality is far easier than recommending a response to it. There are many possible scenarios that might be imagined, from essentially preserving the status quo, to a total front-lines reorganization with regards to MRDF service. What is obvious at Rutgers is that there is an essential need for greater coordination between LS and CS, and among the library units, should we choose to preserve the existing infrastructure. Again, the creation of the MRDF Coordinating Committee will play a key role in achieving smoother communications and a definition of roles among interested parties.

There is however, a national, in fact international, trend which might inform our decision making about infrastructure. The current trend that is being practiced by many research libraries is to concentrate data services within the library, rather than in computer centers. This structure allows for central administrative coordination of the various library aspects related to MRDF: collection development and management, acquisition, bibliographic control, and reference service. There are many libraries which have adopted some form of this type of structure -- the Universities of Michigan, Florida, California at San Diego, Mann Library at Cornell,

Yale, SUNY at Binghamton -- just to name a few, after whom we may choose to model our future infrastructure.

Restructuring to concentrate MRDF services within the library, if done well, could solve many of the problems we currently have in coordinating data services. Such restructuring would minimally require the appointment of a coordinator of MRDF services who would monitor the acquisition, collection development, bibliographic control, reference service, professional training, and bibliographic instruction related to MRDF. Any reorganization would need to be done in recognition of the following:

--Personnel: The libraries cannot absorb the library-type function currently handled by CS without additional personnel having the proper skills. The infusion of GPO-distributed data into our collections has already stretched our ability to provide basic service: we cannot take on more with present staffing configurations.

--A reorganization to concentrate services within the libraries must protect existing services, especially in those areas that are currently well served within the existing structure.

--There will still be a need for close communication and concerted delivery of service between LS and CS.

Recommendations:

Through the MRDF Coordinating Committee, a plan to implement the provision of minimum service levels throughout the Rutgers campuses should be devised.

A subcommittee should be formed to investigate the possible structures for future delivery of services to MRDF, and to author a long-term plan. We should investigate the practice of centralization of MRDF services in other universities and the applicability of these models to RU.

The Task Force recommends administrative consideration of the hiring of additional personnel, or reassignment of personnel, to meet existing needs and provide enhanced support for MRDF.

C. IMMEDIATE CONCERNS:

GPO (especially Census) electronic products are a matter of immediate and grave concern. Census data duplication is and will be rampant. Traditional paper products will continue at many locations as in the past. Electronic redundancy (compact disk products through depository, Rutgers/Princeton Census project, NJ State Data Center and affiliates, bulletin boards, etc.) is a major concern. Furthermore, since Rutgers houses five separate United States federal depository collections, there is a danger of excessive duplication of GPO-produced data throughout the library system.

Collection issues pose a major concern; service issues pose another. Providing service to these products is time-consuming, and requires

computer or database management skills that may not currently be available in all units. LS is not prepared to handle the number of CDs expected to be disseminated over the next few years -- either physically or in terms of service. Quick action is needed.

Recommendation:

Immediate creation of a sub-committee to tackle the pressing problems posed by the GPO electronic data distribution practices. This sub-committee must have representation from the technical side of CS, to assist in reaching technical solutions to problems of access to data, eg. determining physical housing and access to the proliferating GPO compact disks.

II. COLLECTION DEVELOPMENT FOR MACHINE-READABLE DATA FILES

A. AN OVERVIEW

The selection criteria outlined in "Draft Collection Development Statement" (Appendix A) appear eminently practical. However, we also agree that we needed to explore how to broaden the applicability of these criteria from the sole province of Computing Services, to allow for participation in the selection process by the Library system as a whole. With this objective in mind, the following concerns need to be addressed:

--How can we determine what is needed by the research communities that we serve?

--What mechanism or organizational procedures should be developed for implementation of purchase recommendations?

--Who should do the selecting?

--For site specific MRDF (e.g., CD-ROM's or PC products), in the absence of networking capabilities, how can we avoid duplicate purchases of potentially costly products?

--What level of duplication between MRDF products and corresponding paper products is justifiable or desirable?

--Budgetary considerations.

In addition to collection development, collection management and preservation policies are also needed to deal with issues such as the following, e.g.,

--Tapes have a limited lifetime (because of physical deterioration, wear and tear in use, etc.). A maximum of 10 years is the current estimate. Therefore, backup tapes will need to be available, or provision must be made for speedy replacement of failed tapes.

--Similar considerations must be taken into account when dealing with CD-ROM's or PC-diskette products.

--Data obsolescence is another question that falls under the purview

of collection management, and periodic weeding will need to be considered.

--Possible reformatting of data from one format (eg. tape) to another (eg. cartridge or laser disk) is anticipated and must be addressed,

B. OUTLINE OF A SELECTION PROCESS

1. Initiation of Acquisition Recommendations

As a starting point for the consideration of at least of some of the issues raised in the preceding section, we suggest utilization of a selection model already well established in our Rutgers Library system, namely, the subject bibliographers or selectors. These individuals are already knowledgable about specific subject areas. They are also charged with the responsibility for maintaining contact with faculty members (or academic departments) who are active in their subject. These subject bibliographers are therefore also in an excellent position to keep abreast of new developments in MRDF in their respective subject areas and to make recommendations for appropriate acquisitions. From the standpoint of the subject bibliographer, the format of the item recommended for acquisition is less important than the relevance of the contents to the needs of our faculty and students. Viewed in this light, MRDF might just as well be called electronic books.

However, specifically with regard to the selection of MRDF products, this model, based on the use of subject bibliographers must be expanded and modified to fit some new complexities:

a. Provision must be made to allow for MRDF acquisition recommendations originating with people in our system who do not normally function as subject bibliographers. Some of these individuals possess a high level of MRDF expertise and we need a way to ensure that their ideas will be heard. More specifically, we should strive for a spectrum of representation that encompasses both Computer Services personnel and Librarians. A dynamic balance and interaction between these two sister organizations is considered essential. It would also be beneficial if contact could be established and maintained with campus user groups comprising faculty and student researchers.

b. If the position of MRDF Coordinator is established, that individual will, of course, play an active role in the selection process.

Some useful and important functions of the MRDF Coordinating Committee, with regard to collection development and management, might include the following:

--Create awareness of the availability of new MRDF products of possible interest at Rutgers.

--Where MRDF products duplicate paper products, consider to what extent such duplication is desirable or necessary for ease of patron access, and in consideration of the spectrum of user skills and

sophistication (or lack thereof).

--Consider what products ought to be acquired in advance of demand, vs. waiting for specific requests from the academic community.

--Maintain awareness of public domain software and offer recommendations about acquiring those products that would facilitate user access.

--Hardware questions (e.g., computer workstations) will probably also need to be addressed, in conjunction with future acquisitions of other CD-ROM or PC products.

2. What Types of MRDF Products Should be Acquired?

As noted in the "Collection Profile" (Appendix B) the existing collection of tapes in Computing Services is primarily of interest to, and serves the needs of researchers working in the social sciences (e.g., economics, sociology, urban planning, etc.), health care, or business, and usage statistics indicate an ongoing demand for these categories of materials. Because Rutgers also houses five federal depositories, we have a continuing obligation to select appropriate GPO MRDF materials, such as census data, to name one example. Indeed, the volume of such MRDF products offered by the Government is increasing so dramatically that they pose a potential risk of overwhelming recipient libraries with an embarrassment of riches! Also, at the Library of Science and Medicine, we have observed the immediate and enthusiastic patron acceptance of the patent CD-ROM system acquired last year, and the heavy demands made on it both by Rutgers personnel and the community-at-large.

Looking beyond these examples we need to expand our thinking to cover a much broader spectrum of products. Specific areas that have been mentioned by Task Force members as worthy of additional consideration are listed below:

a. Enhanced Products

The Federal Government depository MRDF (mainly CD-ROMs are free, and often come with public domain software (some of the Census CD-ROM's, for example, came with EXTRACT software), usually allowing for basic data retrieval. In order to ensure the most broadly based applications of some of the Government MRDF, collection development activities must include seeking out and evaluating for purchase "enhanced data packages." Enhanced data packages frequently afford user-friendly access to Government data, which otherwise might be difficult to utilize. These packages provide an interface between the user and the data, making the data accessible without any programming expertise. For example, commercial vendors sell Department of State information on CD-ROM's with user friendly software, which otherwise is available on magnetic tape and requires sophisticated manipulation to access. The commercial CD-ROM version allows access to the information by a broader audience which is less sophisticated about the use of computers. Products like these should be considered for purchase at Rutgers. The price of purchase for enhanced data may well be offset by the relief from the necessity of

providing intermediary assistance from library or computing staff.

b. Humanities Products

At present the libraries have an insignificant number of humanities full-text data files This is expected to change with the development of the National Center for Machine-Readable Texts in the Humanities to be located at Rutgers in conjunction with Princeton University. The collecting policies for the Center are broad, and those of the libraries should interface with the Center to avoid unnecessary duplication. The libraries may in some cases chose to acquire texts which the Center would only include in its inventory of available texts and vice-versa. Access and public service policies for these texts should be developed in conjunction with the Center's director so that effort is expended by the appropriate group and each can benefit from the other's expertise.

c. Science Products

MRDF collecting activities in the sciences have also lagged at Rutgers. We are currently witnessing rapid growth in the availability of specialized data collections, in MRDF formats, in the field of chemistry, chemical engineering, toxicology, environmental science, etc. These products are variously offered as CD-ROM's, PC-diskettes, magnetic tapes, and some are also accessible online. For example, enormous collections of various types of physical property data are now commercially available covering, for example (in separate products), mass spectral data, X-ray or electron diffraction crystal structure data, infrared spectral data, ultraviolet spectra, carbon-13 nuclear magnetic resonance spectral data, etc., Other MRDF products now offer thermodynamic data, thermophysical property data, chemical safety data, etc. These products comprise a representative but by no means exhaustive list of what is available. In most cases they are based on or grew out of print products that have been available for years. However, they offer scientists and researchers great improvements in data accessibility and in search power because of boolean features, plus speed and ease of data retrieval. Also, some of the newer products provide the added boon of calculational capabilities which allow for property estimation, extrapolation, and curve drawing. We can expect continuing proliferation in new products of this type and we should be prepared to acquire suitable items for use by Rutgers researchers. However, one caveat must be noted with regard to these files: for the most part, they are commercial products, and therefore tend to be far more costly than the social science or Government MRDF databases. Prices typically range from hundreds of dollars to thousands of dollars for such systems.

For a listing of specific scientific data products, see Appendix C.

III. CATALOGING ASPECTS OF MACHINE READABLE FILES

A. AN OVERVIEW

Cataloging of machine-readable data should proceed in a timely fashion. Information, no matter what its format, should be included in IRIS. LS could assume responsibility for the bibliographic control of all machine-readable data within the university. Until now, cataloging of MRDF has not been done on a regular or uniform basis at RUL. CD-ROMs, software packages, and machine-readable texts in the humanities (for the project to establish a national center at Rutgers for such texts) have been cataloged. However, there is a need to assure uniform treatment of MRDF within the IRIS database.

Recommendation:

We have witnessed a growing number of resources becoming available in machine-readable format; some resources are made available in machine-readable format only. LS needs to establish concrete policies and practices to ensure bibliographic access to MRDF. Across-the-board standards guarantee quality and consistency in the long run.

B. MATERIALS TO BE CATALOGED: LIBRARY-OWNED

Library owned and leased MRDF should be cataloged. Ideally, MRDF owned by other departments within the University will also be cataloged. LS must work with other departments within the University to coordinate the sharing of information to create a database that reflects information resources at Rutgers as a whole (and not limited to the libraries). This will involve changing perceptions about ownership and resource sharing, and changing perceptions can be a slow process.The following conditions will be necessary:

1. MRDF should be cataloged by the Special Formats Cataloging Section. The Section has the necessary expertise and experience to catalog MRDF Further, making the cataloging of MRDF the responsibility of one Cataloging Section will guarantee uniform treatment. The exception may be government depository items. Should RUL decide to load the GPO Catalog tapes, bibliographic records for depository CDs will be included.

2. Full-level cataloging will be provided for MRDF, based on AACR2R, MARC format, and existing national practices, such as those used at institutions which have emerged as leaders in the area of MRDF cataloging, such as Library of Congress, University of Michigan, University of Florida, and Penn State University.

3. Description will be based on the actual item when possible. When this is not possible, the description will be based on documentation.

4. MRDF will be cataloged in RLIN's MRDF (machine-readable data files) file. MRDF is less restrictive than Serials, and accommodates both monographic and serial treatment.

5. Library of Congress Classification and Subject Headings will be used (with the exception of GPO materials, which may receive SuDocs classification).

6. Locally assigned subject headings will also be used. LC Subject Headings don't always accurately reflect or describe the contents of such items as ICPSR data files. A list of locally assigned subject headings may be developed through collaboration of information services librarians and the Special Formats Cataloger.

7. MRDF often include substantial accompanying documentation. Description of the accompanying documentation will be included in the physical description (300 field) of the bibliographic record for the MRDF Both items will receive the same call number, but may be assigned to different shelving locations (for example, the MRDF may be shelved in one location, while the documentation may be shelved in REF.).

8. The material type COMPUTER\FILE (RLIN) and COMFIL (IRIS) are presently used to represent MRDF at RUL. If LS undertakes the cataloging of non-library owned MRDF, new location stamps will be necessary for these materials. While it seems that an exact location stamp for MRDF would be ideal since researchers could then have immediate access, electronic files are volatile in the sense that their physical location may frequently change. If a file needs to be copied to a different tape or disk, then the location stamp would be inaccurate. If a tape management system was used on the IBM 3081, files would constantly be moved according to demand (number of times accessed and mounted). Or, there may be a major change to a new computing environment in which case all locations would change. The ideal location stamp will specify who controls, or "owns," the file, and where to go for more assistance.

C. MATERIALS TO BE CATALOGED: NON-LIBRARY OWNED

LS must also consider the cataloging of information sources which are typically outside the realm or RUL collection and cataloging. Some groups of materials which fall into this category are the holdings of RUCS, ICPSR holdings, departmental data holdings, and the items in the GPO Cataloging tapes. Serial cataloging for some MRDF, such as the ICPSR data holdings, which are fluid in nature, must be investigated.

Recommendation: Full-level cataloging for MRDF is essential in order to provide the highest level of access to our users. Local practices (such as local subject headings and location stamps) will be used as necessary to enhance access.

D. THE BIBLIOGRAPHIC RECORD

The contents of the bibliographic record should accurately reflect to the user: format of the item, extent of the item (file size, content, etc.), location and availability of item, and any restrictions on use. While it is difficult to specify a given size for the standard MRDF bibliographic record (records will vary based on variables such as notes, contents notes, summaries, number of subject headings and added entries), it is necessary to decide upon basic elements that should be present in all records.

The following bibliographic fields will be present in all MRDF records:

--040 Cataloging source
(example- 040 NjR$cNjR)

--1XX Main entry for personal, corporate, or meeting name

--245 Title statement

--256 File characteristics
(specifies type of file, size of file, etc.)

--260 Publication, distribution, etc. information

--300 Physical description

--500 General note field
Will be used for such information as source of title (which is required for MRDF cataloging). May also be used to provide information about accompanying documentation.

--505 Contents note
(Will be used as necessary)

--506 Restrictions on access
(If any)

--520 Summary note
Will be used when necessary to clarify the contents of a MRDF
--650 Subject added entry--Topical term
This field is most frequently used to provide LC Subject Headings

--690 Local subject added entry--Topical term

--7XX Added entry for personal, corporate, or meeting name

MRDF records will not be limited to the preceding bibliographic fields. Other fields (such as series or uniform title) will be used as needed. The point is that we should consistently use the previously mentioned bibliographic fields to guarantee uniform treatment and quality of cataloging.

Recommendation:

MRDF cataloging should be handled consistently. Establishing "core elements" for cataloging will guarantee consistency for access, in our bibliographic database, and in developing the collection.

IV. ACCESS TO MACHINE-READABLE DATA FILES

A. AN OVERVIEW

Access to MRDF involves both storage media and computing capability. The data may reside on any one of the following media: 9-track magnetic tape (round tape), tape cartridge (square tape), CD-ROM, or diskette. The computer used to manipulate the data may be a

mainframe such as the IBM 3081, a minicomputer such as the VaxCluster, or a Sun workstation, which would be connected to the campus network (RUNet). To communicate with this host computer where the processing is being done, the user might use a VT-100 terminal or a microcomputer with communications software running in terminal emulation mode. In cases where the data is locally available on a CD-ROM or diskette, a microcomputer may serve to both access and process the data.

Thus, users of MRDF can be differentiated by their method of physical access to the data, i.e., whether they use a microcomputer to extract data locally, or a terminal to access data remotely. Some users in the second group may elect to download a portion of the extracted data for further processing or manipulation on a microcomputer.

B. THE CURRENT SITUATION AT RUTGERS

Several trends can be identified: users with access to local data files are discovering resources available remotely [PC users FTPing files from data archives]; users with remote access to data want to use files that are currently only available locally [faculty response to CD-ROMs in libraries]; data files are proliferating, especially in CD-ROM format [expect Population Census to come on 3000 discs]; personnel at both RUL and RUCS are not fully meeting the needs of data users; there is a lack of coordination in planning for resources on the campus network (RUNet); users generally are unaware of the vast array of MRDF at RU's numerous locations.

C. SOME TECHNICAL PROBLEMS

Apart from the growing number of users needing to be served and the simultaneous growth in MRDF at RU, there are several technical problems related to computing hardware and storage media. Among these are: the differing "modes" or formats used to store the data - ASCII, EBCDIC, dBase, etc. - which limit usage of the data to a specific computing platform (hardware and operating system); the variety of computing platforms used to access and/or manipulate the data; the need to archive or backup magnetic tapes [or is this mainly a logistical/staffing problem?]; and data files on CD-ROMs or diskettes at RUL that are unavailable or inaccessible via the RUNet.

Recommendations:

a. Long-term goal:

Provide access to data files in all formats to users from all computing platforms.

Actively pursue the development of the RUNet, taking into consideration the many and varied data files available at RU; in other words, make as much data as possible accessible over the network.

Consider adopting new rewritable optical storage technologies such as magneto-optical discs as replacement for existing tape systems, CD-ROMs, and magnetic storage.

b. Short-term goal:

Develop criteria to determine appropriate mode of physical access, eg. remote access, network, stand-alone systems, LANs.

Maintain two levels of access, remote and local, by matching access to format.

Ensure sufficient staffing at RUCS User Services to maintain data tape collection - storing, indexing, cleaning, verification, backup.

Provide adequate online documentation to support remote access to data files, by including codebooks, data dictionaries, variables list, and bibliographic citations.

Acquire appropriate computing hardware to support access to local resources at specific service locations e.g. CD-ROM LANs with jukeboxes or multi-drive units for GPO CD-ROMs at the depository libraries.

Establish/maintain and equip service locations for data access with computing and telecommunications hardware, communications and data analysis software, MRDF documentation, and trained staff.

Determine specific (or types of) local data files to be made available over the RUNet, through the proposed MRDF Coordinating Committee.

V. TRAINING ISSUES

Training of both staff and patrons will closely correlate with the level of service that we determine to provide. Again, setting a minimum level of expectation will enable us to begin the process: advances beyond this level may occur locally and incrementally.

A. STAFF TRAINING

Some skills which librarians will need are already will within their repertoire; others will require enhancements to traditional skills. Depending on the ultimate level of service provision, we may be faced with learning new skills, or hiring personnel already possessing those skills.

The skills which librarians will need in order to provide minimum-level service to machine-readable information are enumerated in the "Levels of Service" portion of this report. Briefly, they are:

--Recognize the need for machine-readable information during the reference interview.
--Knowledge of bibliographic and online finding aids
--Ability of subject specialists and catalogers to incorporate machine-readable sources
--Be able to refer users to subject or computing experts

Training librarians to meet these minimum objectives can occur in a variety of ways: through written guides, workshops, and communication with data users. Additionally, external training is available through the Census Bureau, ICPSR, Association of Public Data Users (APDU), and specialized courses, such as the University of Alberta's summer program in numeric data management. Product-specific training, such as the Patent Office training courses for using the Patent CD-ROMs, is available and should be utilized.

B. USER TRAINING

Again, service level will determine content of any training program. At the least, clear written guides which outline the treatment of data and services offered at Rutgers must be provided. We may wish to embark on an active instructional program for promoting the use of data on campus. This may entail the concerted effort of LS, CS, and interested faculty.

Recommendation:

Obtaining the proper skills to service machine-readable data at RUL is the very crux of the problem. The MRDF Coordination Committee should first consider the issue of librarian training. From there, the Committee can begin to build a user program, drawing on librarians throughout the system who have interest or expertise in data service. Ultimately, a broad-based user training program should be devised.

APPENDIX A

DRAFT

Rutgers University
Computing Services - User Services Division

COLLECTION DEVELOPMENT STATEMENT: POLICY FOR MACHINE-READABLE DATA FILES

INTRODUCTION

This proposed policy is intended to provide guidelines for the acquisition of machine readable data files by the User Services Division of Rutgers University Computing Services. As the technology for the electronic storage of information changes, and as different kinds of information are made available in machine readable formats, this policy may be subject to change. These guidelines pertain to a limited collection of specialized files and new guidelines may be required as the collection grows in size, scope, and format.

DEFINITION OF MACHINE READABLE DATA FILES

A machine readable data file (MRDF) is any information or data, whether numeric, textual, bibliographic, or some combination of these, that is stored in an electronic medium and which is readable only by machine. A typical MRDF consists of numeric data stored as electronically recorded signals on magnetic recording tape and readable by computer. However, other formats, including floppy disks, video disks, hard disks, and compact disks are in existence and may become more common.

PURPOSE OF THIS STATEMENT

The purposes of this machine readable data file policy statement are to:

a) affirm that User Services will acquire information in the format or formats most useful to the Rutgers community;
b) provide criteria for selection of MRDFs including those criteria which are specific only to machine readable data files;
c) outline the role of User Services in providing access to MRDFs.

SELECTION CRITERIA FOR DATA ACQUISITION

Machine readable data files acquired by Information Services must meet several criteria, some of which are the same criteria that are applied to library acquisitions in other formats. However, some criteria are specific to MRDFs. Acquisition of MRDFs by User Services involves the following considerations:

1. Curriculum and Research Support. The machine readable data or information must support an identifiable current or future curriculum or research need that would justify the resources expended in acquiring, processing, and maintaining the files. These expenses include the purchase of tapes, postage, computer time, and staff time to maintain the tapes and document the files.

2. Codebooks and Documentation. In order for a machine readable data file to be acquired, processed and maintained, it must be accompanied by an accurate and complete codebook or manual which includes relevant details of the format and data structure, defines each data element, and includes explanations (dictionaries) of all coding used.

Other documentation, i.e., a description of how the data were collected or a copy of the survey questionnaire, may be required for some data sets in order for the data user to evaluate and use the data.

3. Physical Format. The machine readable format must be compatible with machines (hardware) available to the Rutgers community, specifically, with machines (hardware) maintained by Computing Services.

4. Software. User Services must be assured before acquiring a machine readable data file that the Rutgers community will have the necessary software to access the data file.

5. Duplication of Data. As with other formats, the content of the data should be evaluated in terms of whether it duplicates data already in the the collection. However, it may sometimes be appropriate to make data available in more than one format (e.g. print and machine readable codebooks). Each case will be evaluated by User Services on an individual basis.

6. Authority of Data. As with other formats, the authority and completeness of the data will be considered by the User Services MRDF coordinator.

7. Online Availability. Before acquiring, processing, and maintaining a machine readable data file which is available online through vendors or search services, the feasibility of on-campus availability should be evaluated on the basis of costs and needs. This will be determined by the MRDF coordinator in consultation with representatives from the libraries.

8. Accessibility. Data files that have restrictions regarding their use with which User Services cannot comply will not be acquired. Examples of such restrictions may be that the data may not be copied, or the data may contain proprietary or confidential information.

PROCEDURE

User Services will acquire only those data files that meet the policy criteria stated above. This will normally occur only after a member of the faculty make a request for a specific data file to the MRDF coordinator, but may also occur when a graduate student requests a study for research purposes. Machine readable data files may be acquired for undergraduate use, but only when requested by a faculty member. User Services staff will evaluate the acquisition using the criteria listed above.

User Services will acquire codebooks, user guides, and other documentation along with the data files. These materials will be maintained by User Services and located in the Computing Services Information Center, Room 128, Hill Center. Additional copies of codebooks will also be cataloged and maintained by the Rutgers Libraries.

THE ROLE OF INFORMATION SERVICES

Rutgers University Information Services, specifically the User Services Division of Computing Services, and the libraries, have many roles in assuring campus access to machine readable data files.

1. User Services and the Rutgers University libraries will work together to identify the existence of data files not in library collection and to determine if they can be acquired. Library and User Services staff will continue to actively inform users of data pertinent to their needs and to identify machine readable data.

2. User Services will be responsible for the acquisition of data files and for insuring that the files meet the standards set forth in the selection criteria listed above. When data files are available from more than one source, User Services will determine the most reliable and useful source.

3. Since Computing Services is a Primary Participant in the New Jersey State Data Center, and the Library is a federal government depository library, User Services has a commitment to making government-produced information available at Rutgers and to ensuring access to government-produced information in electronic formats.

4. User Services will coordinate memberships in organizations such as the Inter-University Consortium for Political and Social Research (ICPSR), the New Jersey State Data Center, the Association of Public Data Users (APDU), and the Roper Center.

5. If funds for a specific request for MRDFs are not available, User Services will attempt to coordinate funding with participating departments, as is currently done to acquire Compustat datasets from Standard and Poor's Corporation.

6. User Services and the Library will acquire, maintain, and catalog codebooks and other MRDF documentation, making them readily available to users.

7. User Services and the Library will provide reference service for identifying machine readable files in the Computing Services collection. Referrals to machine readable files will be provided as a basic part of reference services.

8. User Services will provide a further level of technical services to users about accessing data files and tapes, including provision of information regarding specific tapes and files, statistical packages, job control language, and other relevant assistance such as seminars and online documentation.

9. Computing Services will physically maintain machine readable data files in the forms of tapes, compact disks, floppy disks, and other electronic mediums.

10. User Services will produce guides, catalogs and other materials in print and machine readable format to aid in identifying and locating machine readable files in the collection.

11. User Services will participate in organizations pertinent to MRDFs, including the International Association of Social Science and Information Technology (IASSIST) and the Association of Public Data Users (APDU).

12. User Services will continue to facilitate communication with and between the many departments that use MRDFs with regular meetings of the Data Base Advisory Committee, which is composed of representatives of Information Services and faculty from each campus, including Camden, New Brunswick, and Newark.

13. User Services will continue its cooperative relationship with Princeton University in the Princeton-Rutgers Census Data Project and the Princeton-Rutgers joint membership in the Roper Center.

Mary Jane Cedar Face

DRAFT

Rutgers University
Computing Services - User Services Division

COLLECTION PROFILE: MACHINE READABLE DATA FILES

Subject: Machine Readable Data Files (MRDF)

Bibliographer: Mary Jane Cedar Face

Number of Departments Served: Over 30 departments.

Collection Size: The collection, as of the Fall, 1990, includes approximately 500 machine readable data studies consisting of over 6000 separate computer files (datasets) on over 700 reels of computer tape. The number of tapes includes both working copies and backup copies. In addition to the machine readable data files, the collection includes approximately 500 unique volumes of printed codebooks and other documentation that accompany the datasets, with additional copies of many codebooks.

Locations: The MRDF collection is maintained by the User Services Division of Rutgers University Computing Services. Data files on tapes account for the vast majority of data files and are stored in the Operations Area in the basement of Hill Center, Busch Campus. (Although the tapes are physically stored at the Hill Center, they are accessible through terminal servers from on and off campus to researchers with accounts on the IBM 3081 system.) Diskettes with MRDFs are located in the Microcomputer Resource Center in Hill Center and CD-Roms are currently located in Hill Center 128. Codebooks to the data files are housed in the Computing Services Information Center, Hill Center 128, and additional copies of many codebooks can be found on the College Avenue Campus in Alexander Library and on Busch Campus in the Library of Science and Medicine. Documentation to the Compustat datasets is located in several additional departments.

Scope: Machine Readable Data Files (MRDFs) constitute a collection of data that is acquired or stored in a format which requires a computer to read or use the information. This collection profile covers numeric data and text files used primarily for social science and business research and instruction. This profile does not cover computer-stored graphic images or computer programs.

Purpose: The purpose of the MRDF collection is to support research and instruction in those Rutgers University programs which require information in machine readable form and to collect that information which is both available only in machine readable form and needed for research or instruction at Rutgers University.

At present, the collection of machine readable data files supports graduate and faculty research in over thirty academic departments, primarily the departments of Sociology, Urban Planning, Economics, Political Science, Business, and Social Work. The Institute for Health, Health Care Policy, and Aging Research and the Graduate School of Management also make extensive use of this collection. Undergraduates in the Social Sciences (primarily Economics, Political Science, and Sociology) also use the data files in their academic programs.

The collection of codebooks provides information about the machine readable data files. The codebooks and other accompanying documentation provide necessary information for utilizing the computer files.

Collection Guidelines:

A. Physical acquisition of MRDFs and documentation:

Currently, magnetic tape is the primary medium for acquisition of machine readable information. The normal procedure is for User Services staff to send blank magnetic tapes to the appropriate data center with preassigned tape numbers, so that the selected datasets can be copied onto them. In addition, User Services acquires some MRDFs in other formats such as personal computer floppy discs, compact discs, and by electronic transfer (e.g., electronic mail and downloading from bulletin boards).

User Services acquires the necessary datafile documentation (e.g., codebooks, user guides, technical documentation) in hardcopy and/or machine readable format. The particular format chosen depends on availability, (some documentation is available in only one physical format), ease of use, and size and cost (some documentation is so voluminous that printed copies are very expensive). When datasets are ordered from the Inter-university Consortium for Political and Social Research (ICPSR), an additional set of documentation is also automatically ordered for the Rutgers University libraries.

B. Storage of MRDFs and documentation:

Most datafiles are stored on magnetic tapes by Computing Services in the Hill Center basement on Busch Campus. The tapes are 2,400 feet long, with a density of 6250 bits per inch, each of which could hold a maximum of up to 170 million bytes of data. The tapes are generally formatted with standard labels that can be mounted on the IBM system. Researchers with IBM accounts can access these tapes with "read only" privileges. Relevant data can then be extracted and copied to researcher accounts or departmental systems, or downloaded to personal computers for secondary analysis. There is also a small collection of diskettes that contain numeric data. These are located in the Information Center, Hill Center, Room 128. Some data is beginning to be acquired in compact disc format.

User Services has an ongoing program of copying selected files from older tapes to new tapes. Backups are not routinely made of all files due to storage considerations and the ease of acquiring replacements from ICPSR.

Print copies of codebooks and other relevant documentation are kept in the Information Center of Hill Center. Machine-readable codebooks are stored on tape with the datafile being documented. Additional, circulating, copies of many codebooks can be found in Alexander Library and the Library of Science and Medicine.

C. General Considerations

For the most part the MRDF collection is being built passively in response to requests for machine readable data from faculty and graduate students. There is no active program for acquiring data in anticipation of possible future needs. The exception to this general rule is the acquisition of new releases of a few heavily used studies as they are made available. Examples include: The General Social Survey, The American National Election Study, the Survey of Income and Program Participation, and selected census data files.

The MRDF collection consists almost exclusively of numeric data. However, some textual data is being acquired, notably codebooks and other documentation that can be printed offline. Although this collection profile reflects the needs of social science research, it is possible to modify this profile to include other areas and more textual material.

The large majority of data studies acquired so far by User Services have been acquired from the Inter-University Consortium for Political and Social Research (ICPSR). The Library pays an annual fee to the Consortium in order to maintain the university membership with the Consortium. The MRDF bibliographer is the Official Representative of Rutgers University to ICPSR. Rutgers University also shares a joint membership with Princeton University in the Roper Center for Public Opinion Research. The Rutgers portion of the Roper Center membership fees is paid by the library, and User Services regularly acquires data files from this source. Additionally, Computing Services is a primary participant of the New Jersey State Data Center, through which census data is acquired by User Services staff. User Services coordinates the funding for the annual acquisition of the Compustat datasets from Standard and Poor's Corporation, which contain financial, statistical and market information for several thousand companies. The remaining data files have been acquired from miscellaneous sources.

D. Available resources

The Association of Public Data Users (APDU) was organized in 1976 to encourage interaction and the sharing of information among data users, producers, and distributors.

The International Association for Social Science Information Service and Technology (IASSIST) is an international organization of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable data text and/or numeric social science data..

The Inter-University Consortium for Political and Social Research (ICPSR) is the primary archive for machine-readable social science data in the United States, and the largest social science data archive in the world.

The Roper Public Opinion Research Center at the University of Connecticut is an archive of sample survey data from over seventy countries.

The New Jersey State Data Center, established in 1978 as part of the State Data Center Program of the Census Bureau, receives census data and disseminates this data to the public through a network of affiliate agencies around the state. Rutgers University is a primary participant in this program.

Mary Jane Cedar Face

APPENDIX C

SCIENTIFIC DATA SOURCES

NAME	DESCRIPTION
NIST/EPA/MSDC Mass Spectral Database	Spectra of 48,000+ compounds; accessible by name, or CAS registry no., or molecular formula, or molecular weight, or via spectral characteristics.
Powder Diffraction Data Files from ICDD (Internat. Center for Diffraction Data)	X-ray or electron diffraction crystal structure data for tens of thousands of chemical compositions.
Sadtler Spectra	Infrared spectra, carbon-13 NMR spectra, UV spectra, etc.; over 120,000 spectra listed.
TDS Numerica	Offers a family of chemical & physical property databases covering a variety of research areas, e.g., -DETHERM: thermophysical data -CHEMSAFE: chem. safety data -PPDS: thermodynamic data -CHEMEST: chem. property estimation procedures. -TRC: vapor pressure data -ENVIRONMENTAL FATE DATABASE -ACQUIRE: aquatic toxicity data -CIDES: carcinogenicity data etc.